Towards an optimal use of numerical precision in Earth System models: the case of NEMO

Oriol Tintó Prims[1,2], Mario Acosta[1], Miguel Castrillo[1], Kim Serradell[1], Alícia Sanchez[1], Ana Cortés[2], Francisco Doblas-Reyes[1,3]

[1]Barcelona Supercomputing Center , [2]Universitat Autònoma de Barcelona, [3]Institut Català de Recerca i Estudis Avançats

Using more numerical precision than needed limits our capability to perform bigger and more complex experiments. Adapting Earth System models to use just the necessary precision will make them faster and cheaper!

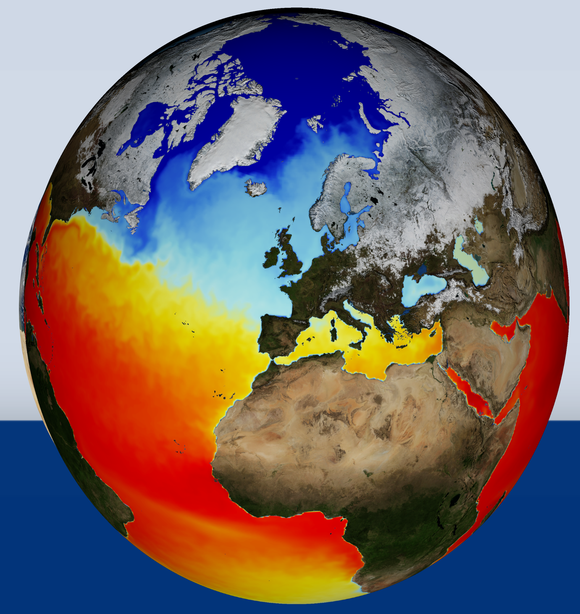
What is NEMO?

NEMO is an global ocean circulation model that includes several sub-models:

- OPA, the nucleus, solves the dynamics and thermodynamics of the ocean.
- LIM, a sea-ice model.
- TOP, a biochemistry model.

Why is NEMO important?

NEMO is developed by an European Consortum and used by hundreds of institutions. In the CMIP6 project many coupled models use NEMO as its oceanic component. Several billions of computing hours will be invested in these simulations.



Autosubmit

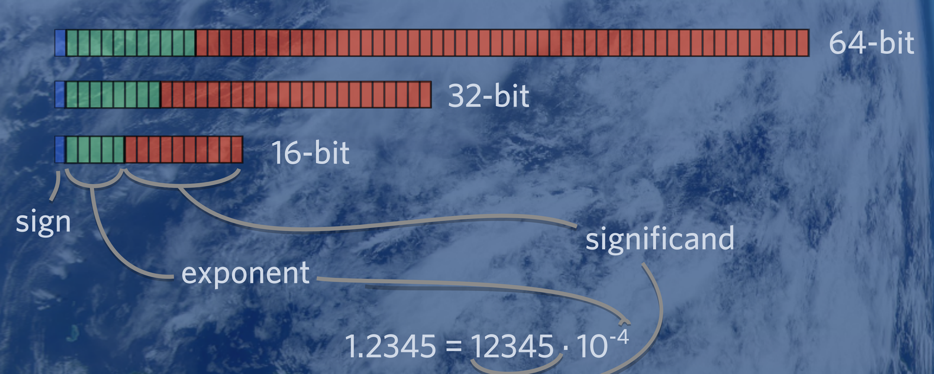
Autosubmit is a python-based tool to create, manage and monitor experiments by using Computing Clusters, HPC's and Supercomputers remotely via ssh. It has support for experiments running in more than one HPC and for different workflow configurations.

Reduced Precision Emulator

The **Reduced Precision Emulator**[1] is a library which has the capability to emulate the use of arbitrary reduced floating-point precision within large numerical models written in Fortran.

[1]Dawson, A. and Düben, P. D.: rpe v5: an emulator for reduced floating-point precision in large numerical simulations, Geosci. Model Dev., 10, 2221-2230, https://doi.org/10.5194/gmd-10-2221-2017, 2017.

Why is it worth to explore mixed-precision approaches?



The use of mixed-precision approaches helps to better use computational resources, reducing the memory usage, the data movement and better exploiting vectorization.

In Earth System models, there is room to reduce the precision used, adjusting to the real model requeriments and leaving behind the bad habit of over-engineering this resource.

The adoption of mixed-precision approaches by the state-of-the-art Earth System models models will represent a huge saving of computational resources.

Making the adaptation is not simple and there's a need to make the process easier.

To do so, we developed a methodology to obtain a mixed-precision version of a full-scale model minimizing the user intervention.

About the methodology

Implementing the emulator

Even the **reduced precision emulator** it is a valuable tool to explore reduced precision approaches, its implementation to a complex model is not trivial.

To facilitate the process, a Python tool has been developed to parse the model sources and automatically modify the code to implement it, minimizing user intervention.

Evaluating the error

Evaluating if the error is acceptable or not is problem dependant. The usual approach consists in comparing a set of output variables against a reference double-precision simulation of the same problem, using a scalar magnitude like the Root-Mean-Squared Error.

Designing the tests

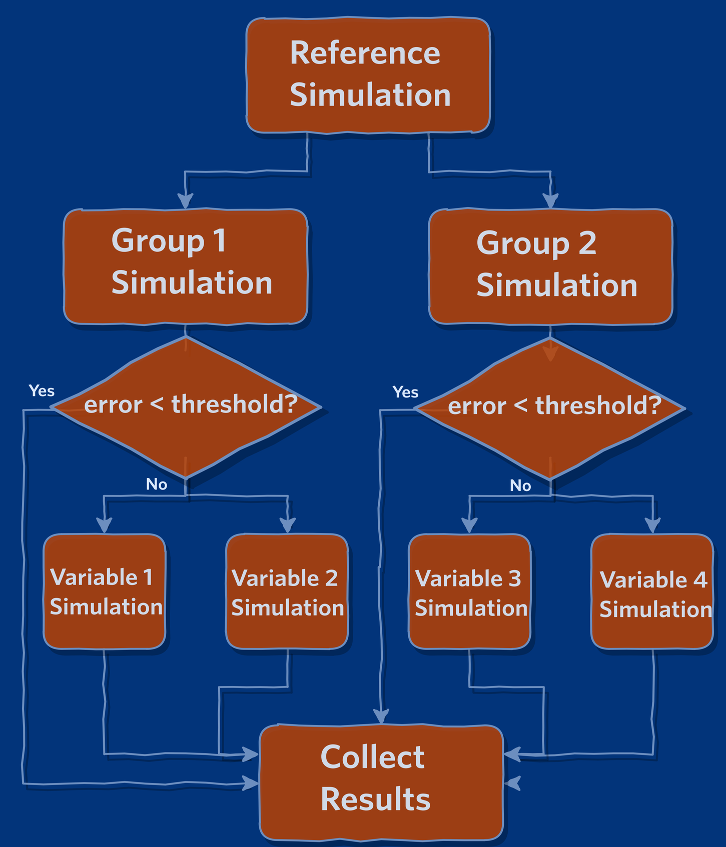
State-of-the-art Earth System codes use thousands of real variables. To reduce the number of tests that are necessary to screen all the variables we adopted a grouping strategy. The variables are tested in groups in the following way: the precision of all the variables in a group is reduced to the desired precision; if the error resulting from the simulation is smaller than our defined threshold, we consider it safe to reduce the precision of all those variables, if otherwise the error is bigger, the variables of the group are tested individually.

Managing the simulations

With the strategy adopted to design the tests, we will obtain a workflow consisting of many group simulations, and depending on the output of each one of those simulations, the test of all the individual variables belonging to a group.

To manage all those simulations and its dependencies, we used the Autosubmit workflow manager.

Using this tool, all the process of testing and processing the results is automatic and does not require user intervention.



Environment

Marenostrum IV

- **@ Barcelona Supercomputing Center**
- Peak power is **11.15 Petaflops.**
- **165,888 CPU cores** in 3,456 nodes.
  - 48 CPU cores per node.
  - 96 GB of main memory per node.

Proof of concept

Model and version:

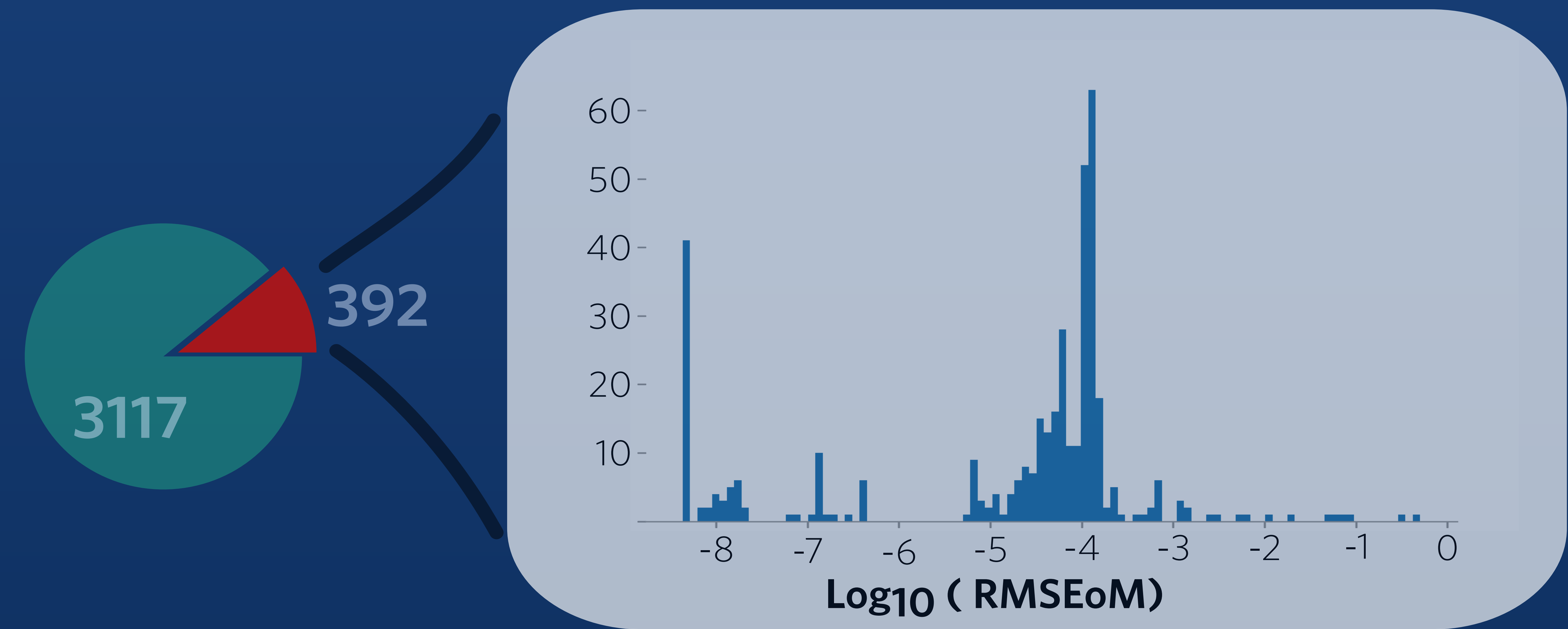
- NEMO v3.6 Stable

Configuration:

- GYRE 1 (32x22x10)

Diagnostic based on:

- Root-Mean-Square Error over Mean of the column heat content



The proposed methodology has been tested using the NEMO model, with a GYRE configuration. Setting an error threshold of 10^{-10}, the final number of tests was 1400 to screen 3509 variables.

The figure shows that from all those variables, 89% can safely use a 10-bit significand, proving the huge room that exist for reducing the precision in this model.

Among the 392 variables that produced a bigger error than the threshold, a large proportion were producing an error of magnitudes between 10^{-3} and 10^{-5}, that probably depending on the specific application can be perfectly tolerated.

Aside from NEMO, this methodology and tools have ben tested using the ROMS model for a 4D-VAR data assimilation problem, showing promising results too.

Future Work

- Develop the tool to automatically assimilate the results and build an actual mixed precision implementation that can preserve double-precision quality.
- Evaluate the impact in the computational performance.
- Ensure compatibilitiy with other models.
- Document the tools developed and **make it available fo interested users.**

for further information please contact with **oriol.tinto@bsc.es**